Theor Appl Genet (1994) 89: 64–69

Y. X. Cui · G. W. Xu · C. W. Magill · K. F. Schertz
G. E. Hart

A low-copy-number *Sorghum* DNA sequence that detects hypervariable *Eco*RV fragments

Received: 30 October 1993 / Accepted: 12 January 1994

Abstract A sorghum genomic DNA clone that hybridized on Southern blots in simple but different patterns to fragments produced by digestion of DNA from the parents of an F_2 mapping population was hybridized to *Eco*RV-digested DNA from 53 accessions. Forty-six different fragment patterns were observed, each comprised of from one to ten bands. Much less variability was detected in *Eco*RI than *Eco*RV digests of a selected subset of the accessions. Base-sequence analysis of the clone did not reveal a functional identity for the sequence and the clone does not contain repeated sequences often associated with hypervariable loci. Clones such as this will be especially useful in evaluating germplasm diversity and in identifying the potential parentage of hybrids.

Key words Hybridization probe · *Sorghum bicolor* Fingerprinting

Introduction

Knowledge of the origins and phylogenetic relationships of important crop species has practical as well as esoteric value. The origin and phylogeny of the species, subspecies, and races of sorghum, the fifth most important cereal crop on a worldwide basis (Doggett 1988), has been the subject of morphological, cytogenetic, cytoplasmic, isozyme and RFLP analysis (Liang et al. 1966; de Wet 1978; de Wet et al. 1967, 1970; Schertz et al. 1978; Mann et al. 1983; Morden et al. 1989, 1990). Although the results of these studies contributed in complementary ways to our current understanding of the sorghum species complex, they have not always agreed. Sorghum classifications have been modified several times, especially at the level of subspecies and races.

Sorghum bicolor (L.) Moench, which is a highly diverse species, belongs to the genus *Sorghum* of the tribe *Andropogoneae*. It has been divided into three subspecies, namely, *bicolor*, *verticilliflorum* and *drummondii* (Doggett 1988). Subspecies *bicolor* is recognized as consisting of five basic races, bicolor, kafir, caudatum, durra and guinea, and an additional ten intermediate races (Harlan and de Wet 1972). The four races verticilliflorum, arundinaceum, virgatum and aethiopicum are recognized in the subspecies *verticilliflorum* (Doggett 1988).

It has been proposed that modern sorghums have diverse origins, with the cultivated subspecies *bicolor* arising from the wild subspecies *verticilliflorum*. It is thought that the wild race aethiopicum gave rise to the durra and bicolor cultivated races, while the wild races arundinaceum and verticilliflorum gave rise to guinea and kafir types, respectively (Mann et al. 1983). However, de Wet was of the opinion that durra arose from kafir (de Wet et al. 1967, 1970).

Apparently, several factors contributed to the continuum of variant subpopulations in sorghum (Doggett 1988). In addition to the diversity attributed to multiple origins, the high rate of interfertility among sorghum relatives means that crossbreeding can lead to highly-polymorphic populations. Though self-fertilization is the norm, outcrosses between cultivated and highly-variable wild species often found in the same locale are believed to occur frequently and to stimulate gene flow, genetic introgression, and the formation of intermediate types. These processes make the classification of sorghum difficult in practice and the position of accessions between and within races is often very tenuous.

Communicated by A. L. Kahler

Y. X. Cui · C. W. Magill (✉)
Department of Plant Pathology and Microbiology,
Texas A&M University, College Station,
Texas 77843, USA

G. W. Xu · K. F. Schertz · G. E. Hart
Department of Soil and Crop Sciences,
Texas A&M University, College Station,
Texas 77843, USA

K. F. Schertz
USDA-ARS, Texas A&M University,
College Station, Texas 77843, USA

Isozymes have been used to identify unneccessary duplication of accessions within and between the large collections of sorghum germplasm maintained at ICRISAT (International Crops Research Institute for the Semi-Arid Tropics) and in the United States. However, insufficient polymorphism was detected to confidently differentiate between similar accessions (Aldrich et al. 1992). RFLPs have generally proven to be a more powerful tool for monitoring genetic variation. Accessions can be probed with a set of unique-sequence clones and/or repeated-sequence (fingerprinting) clones, differences and similarities determined, and attempts thereby made to maintain a diverse collection of genotypes while avoiding over-representation of closely-similar or identical accessions. The clones can also be used to identify varieties, protect breeders rights, and identify parentage in a breeding program. A low-copy-number sorghum DNA clone, selected for use as a probe in genetic mapping and diversity analysis, detected an unusually high level of polymorphism among *Eco*RV fragments. This clone is the subject of our report.

Materials and methods

Plant materials

Fifty-three sorghum accessions from nine races and two subspecies of *S. bicolor* (L.) Moench were used in this study (Table 1). Young leaves from two to five seedlings of each sorghum accession grown in a greenhouse were collected, the midribs removed, and the tissue was lyophilized, ground to a fine powder and stored in a –20 °C freezer.

DNA extraction, digestion, Southern blotting and RFLP probing

Genomic DNA was extracted from 0.6 g of dried powder (Murray and Thompson 1980; Xu et al. 1993). Aliquots containing 8–10 mg of DNA were digested in separate reactions with the restriction endonucleases *Eco*RI and *Eco*RV. DNA-fragment separation, Southern blotting, probe labelling, hybridization and autoradiography, followed standard procedures (Maniatis et al. 1982; Helentjaris et al. 1986; Reed and Mann 1985; Feinberg and Vogelstein 1983). Two duplicated accessions, independently carried through each step, provided a check for consistency.

Origin and preparation of pSbTXS1075

Plasmid SbTXS1075 contains a 1-kb *Pst*I fragment of sorghum genomic DNA. It is one of many clones identified as suitable for use in mapping based on the observation that simple, polymorphic hybridizing patterns are produced on blots of *Eco*RV-digested DNA from accessions BTx623 and IS3620C, the parents of an F_2 mapping population. Polymorphic fragments detected by the insert in pSbTXS1075 (abbreviated hereafter as TXS1075) have been mapped to sorghum linkage groups D and I (Xu et al. 1993).The insert was removed from the plasmid by *Pst*I-digestion and was isolated in low-melting-point agarose, following electrophoresis for use as a probe.

DNA sequence analysis

The base sequence of TXS1075 was determined using Sanger's dideoxynucleotide chain-terminating method (Sanger and Colson 1975; Sanger et al. 1977) as recommended for Sequenase Version 2.0. The sequences of both DNA strands were completed by primer-walking. The sequence was analyzed using MacVector™ software, Version 3.5. Searches of nucleotide and protein sequences in the GeneBank and EMBL databases were performed at the NCBI using the BLAST network services.

Table 1 Accessions of *Sorghum bicolor* included in this study

Accession	Race	Origin	Lane in Fig. 2
spp. *bicolor*			
Chinese Amber FC8728	bicolor	China	16
Chinese Nat. Acc. No. 422	bicolor	China	1
CI171	bicolor	China	22
EBA-9	bicolor	Sudan	18
IS1598C	bicolor	Indian	26
IS6964	bicolor	Sudan	28
IS7542C	bicolor	Nigeria	27
Shanqui Red	bicolor	China	23
Standard Broomcorn CI556	bicolor	Italy	10
Sweet Sudangrass SA372	bicolor	Sudan	12
IS12568C	caudatum	Sudan	29
IS12608C	caudatum	Ethiopia	30
IS6710C	caudatum	Senegal	31
M91051	caudatum	USA	24
BTx623	Caudatum/kafir	USA	2
3-Dwarf White Sooner Milo	durra	unknown	7
IS1022C	durra	India	32
IS12570C	durra	Sudan	34
IS7333C	durra	Nigeria	33
SA7078	durra	unknown	8
TX403	durra	USA	25
IS3620C	guinea	Nigeria	15
IS5332C	guinea	India	36
IS7173C	guinea	Tanzania	20
IS7920	guinea	Nigeria	9
IS7920C	guinea	Nigeria	35
IS3477C	guinea	Sudan	38
IS3614C	guinea	Nigeria	37
IS3955C	guinea	Nepal	39
IS7419C	guinea	Nigeria	17
87-4325	kafir	S. Africa	40
87-4326	kafir	S. Africa	41
87-4327	kafir	S. Africa	42
87-4328	kafir	S. Africa	43
87-4329	kafir	S. Africa	44
BTx378	kafir	USA	3
BTx398	kafir	USA	4
BTx3197	kafir	USA	5
B35	unknown	USA	21
QL3-India	unknown	India	11
RTx430	unknown	USA	6
spp. *verticilliflorum*			
IS14564	aethiopicum	Sudan	14 & 45
IS18820	aethiopicum	Egypt	47
PI302105	aethiopicum	Ethiopia	46
IS18826	arundinaceum	Ivory Coast	13 & 48
PI156549	arundinaceum	Rhodesia	49
PI186570	arundinaceum	Nigeria	50
IS14505	verticilliflorum	Unganda	52
IS18798	verticilliflorum	S. Africa	51
PI208190	verticilliflorum	unknown	19
PI267331	verticilliflorum	India	53
IS18803	virgatum	unknown	55
IS18809	virgatum	Egypt	54

Results

Race comparisons using TXS1075

Initially, 18 accessions (two representatives from each of the nine races that comprise two subspecies) were studied.

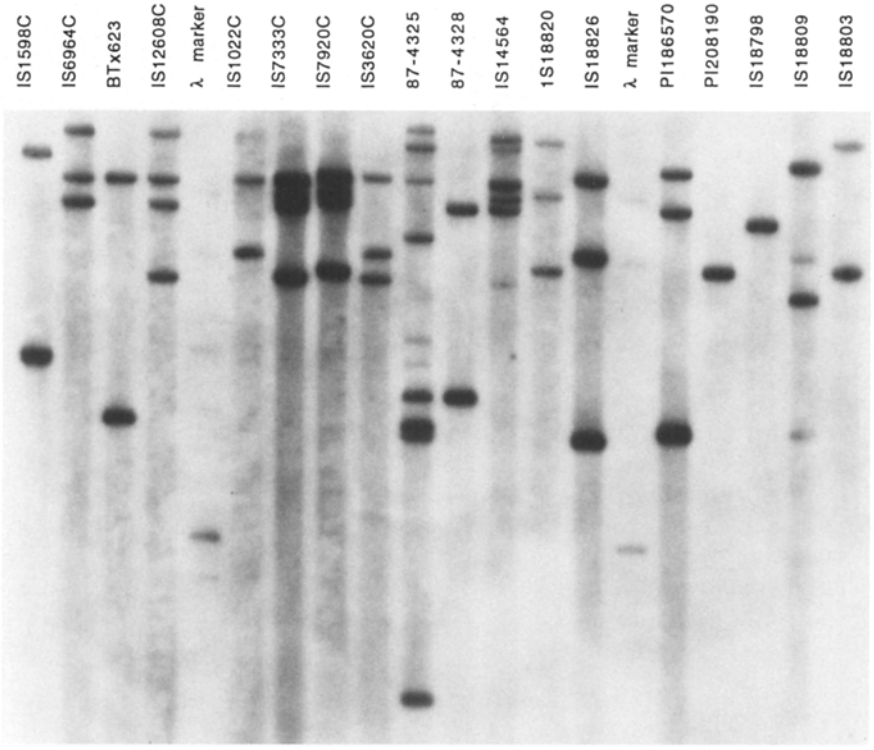

Fig. 1 Southern blot obtained by hybridization of TXS1075 to *Eco*RV-digested DNA from two accessions of each of nine races

Digested DNA samples obtained from accessions of the same race were loaded into adjacent lanes for electrophoresis, except that *Hin*dIII digests of λ DNA were loaded in the 5th and 15th lanes to serve as molecular-weight markers. The pattern of fragments detected following digestion with *Eco*RV revealed >94% polymorphism (Fig. 1); only two of the 18 accessions shared the same pattern. From one (PI208190 and IS18798) to ten (87-4325) hybridizing fragments were detected among the accessions. Some accessions (PI208190 and IS18798, or IS18826 and PI186570) had the same number of fragments but the molecular weights of the fragments were different. Accessions 87-4325 and 87-4328, both classified as kafirs originating from S. Africa, have very similar morphological characters but very different *Eco*RV-fragments; 87-4325 has ten hybridizing fragments, while 87-4328 has only two. Similarly, IS3620C and IS7920C, both classified as guineas originating from Nigeria, have very different RFLP patterns. The latter accession gave the same pattern as IS7333C, which also originated from Nigeria but is classified as a durra.

The results suggested that TXS1075 may be very efficient for distinguishing individual sorghum accessions regardless of the species, subspecies, or race to which the accessions belong. Surprisingly, the degree of polymorphism detected with *Eco*RI digests was much less; only five patterns were detected, and one of these constituted a single 4.3-kb fragment that was present in 11 of the 18 accessions (data not shown).

Analysis of additional accessions with TXS1075

To further explore the potential use of TXS1075, the study was extended to include 35 additional accessions. The hybridization patterns for *Eco*RV digests for the 18 original and the additional 35 accessions are shown in Fig. 2. Forty-six RFLP patterns were observed. Four patterns were shared by two or more accessions. Accessions BTx378, BTx398, BTx3197, 87-4326 and 87-4329 (lanes 3, 4, 5, 41, and 44, respectively, in Fig. 2), all of which are kafirs, display a single 3-kb fragment. BTx378, BTx398 and BTx3197 are parents of hybrids that have been selected in breeding programs, and have shown identical RFLP patterns with a number of probes.The South African accessions 87-4326 and 87-4329 have shown identical RFLP patterns with all probes so far tested (unpublished results) as have the 3-Dwarf White Sooner Milo and Tx403 accessions. Two other pairs of accessions, QL3-India and IS7173C, and IS7920C and IS3614C, which have been shown to be similar, but not identical, in previous tests (unpublished results), were not differentiated by this probe.

Each of the remaining accessions had a unique RFLP pattern. Even BTx623 and BTx3197, related accessions which share many morphological characters, can be distinguished by TXS1075. These results provide further evidence that TXS1075 can distinguish among numerous accessions, whether they belong to the same or different subspecies or races.

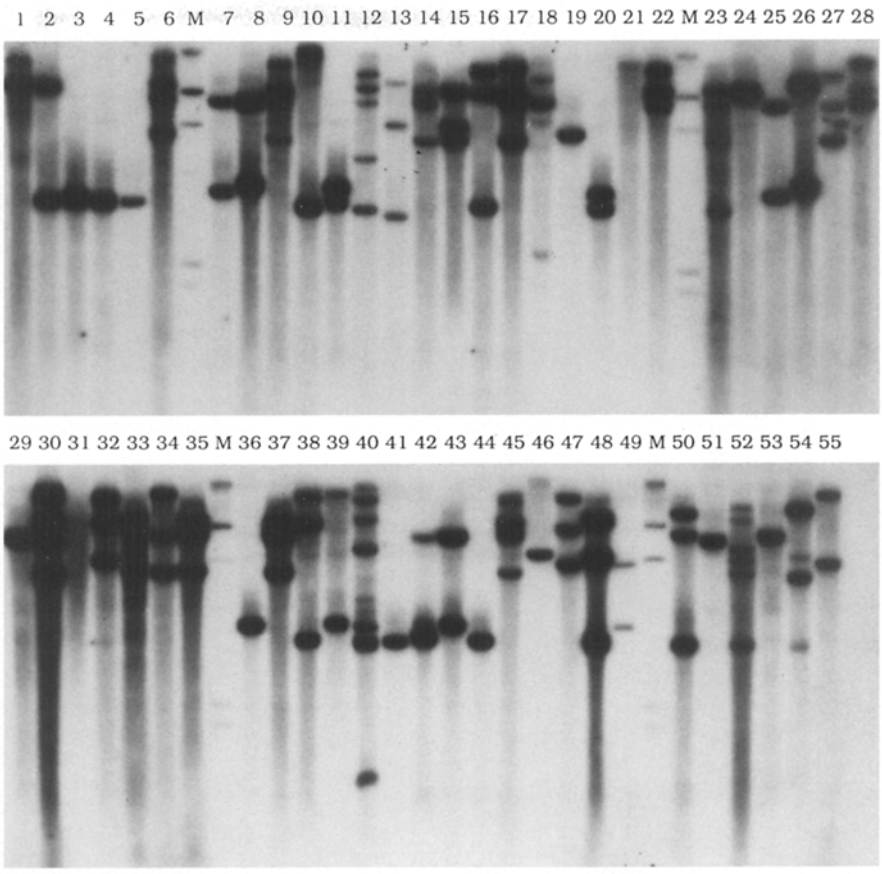

Fig. 2 Southern blot obtained by hybridization of TXS1075 to *Eco*RV-digested DNAs from 53 sorghum accessions. Identity of the accession in each numbered lane is given in Table 1. Lanes marked *M* contain *Hin*dIII-digested *lambda* DNA detected by including a small amount of *lambda* DNA in the labeling reaction

Nucleotide sequence of TXS1075

The complete nucleotide sequence of TXS1075 has been deposited in GenBank (accession no. U02558). The clone is 1014 bp in length. No significant homology to any DNA sequence in GeneBank or EMBL was detected using a BLASTN search (Altschul et al. 1990). A long potential open reading frame includes the first 381 bases and several other shorter ones are present, so a BLASTX search was also made using filters to eliminate regions of low information. The high-scoring matches included an α-lactalbumin, a nucleolar protein, and the reverse transcriptase from cauliflower mosaic virus, with probabilities that the matches would occur by chance of 6.4%, 16.0%, and 23.0%, respectively. Though several potential 5′ and 3′ intron splice sites and poly-A tail signals are present in the sequence, they are not in positions that would suggest functionality for any of the open reading frames. Other features present include a 10-base palindrome starting at base 208, several short, direct repeats (the longest 11 bases in length) and inverted repeats, two of which have perfect matches over a length of 12 bases.

The sequence is A-T rich (59.2% overall and 61.0% within the non-open reading frame regions), which is the average A-T content in introns in monocot plants (White et al. 1992). Dinucleotide group-analysis reflects the overall base composition in that CC and CG pairs are lowest in frequency, while AA, AG, and TT pairs are the most frequent.

Discussion

Quick and simple methods for monitoring the similarities or differences among plant materials have several practical applications. Taxonomists have recognized that overduplication is a significant problem in the storage of plant germplasm for conservation (IBPGR 1984). With limited resources and facilities, it is important that the maximum amount of diversity be maintained. Plant breeders and seed producers would also like to take advantage of tools that identify similarities and differences among breeding stocks or segregating progeny, or that provide tags for protecting breeder's rights. The potential for detecting misclassified lines or errors in labeling of seedlots using TXS1075 as a DNA hybridization probe was demonstrated by the observation that two accessions from Nigeria, IS7333C and IS7920C, classified as a durra and guinea respectively, gave the same *Eco*RV restriction pattern. Results of probing the DNA of these two lines with several other probe and restriction enzyme combinations (unpublished) reveal they are very closely related if not identical.

As for other plant species, early sorghum classification schemes were mostly based on a few morphological characters (for a review see Doggett 1988). Later, isozyme markers were used to study sorghum diversity in accessions collected from various geographic locations or taxonomic groups, including sections, species and subspecies. As in other species (Bernatzky and Tanksley 1989; Lubbers et al. 1991), the amount of variation in isozymes detected among sorghum accessions (Morden et al. 1989, 1990; Aldrich et al. 1992) is low when compared to that for RFLP's (Aldrich and Doebley 1992).

Cultivated and wild sorghums are widely distributed around the world. Without barriers to outcrossing, genes can easily flow between races, subspecies and even species. Morphologically-distinct accessions may in fact be closely related while those with similar appearances may not be. An approach that seems to be quite promising at the species level is the use of probes from repeated sequences. Hoang-Tang et al. (1990) have identified clones that are specific for *S. bicolor* and *S. halepense* or *S. versicolor*. When such a probe reveals numerous distinct DNA fragments, the pattern may also provide an identifying "fingerprint" for an accession. However, for identifying heterozygotes in populations, and for evolutionary studies, the simple patterns seen for unique sequence probes are preferable.

The clone described in this paper, TXS1075, has the remarkable ability to distinguish 46 of 53 accessions on Southern blots of DNA digested with a single restriction enzyme, *Eco*RV. Ten-race bicolor accessions each gave a unique pattern; eight kafir accessions gave four different patterns; six durra accessions resulted in five patterns; nine guinea accessions formed eight patterns, and 12 accessions from the wild subspecies *verticilliflorum* each had a unique pattern. How can the hypervariability seen for this probe be explained? Other than for immunoglobins and T cell-receptor genes in animals, where site-specific recombinases generate sequence diversity, hypervariable loci have generally been associated with tandem repeats. Repeats that are runs of a single base, or of trinucleotides such as those that have recently been found to account for several human diseases with variable expression (Kremer et al. 1991; Brook et al. 1992; Redman et al. 1993), or those that are much longer, such as a VNTR repeat of 26 bases which contains an inverted repeat (Fraser et al. 1989), all seem to generate local instability in the genome. Clone TXS1075 does not contain repeats near the length of those that have been identified as hypervariable; the longest trinucleotide repeat was a run of four GTAs. It also seems unlikely that the clone is adjacent to a repeated sequence, because then only length and not numerical changes would be expected. The same argument can be used for the clone being part of an intron. Though introns do accumulate mutations at a greater rate than coding sequences (Lewin 1990), they would only rarely result in changes in the number of fragments detected. A more likely explantation is that the clone is a segment or remnant of a retrotransposon-like element. This would account not only for the apparent differences in copy number in different genomes, but also for the great difference in polymorphisms detected by two restriction enzymes. If one *Eco*RV site is in the element and the second in the flanking DNA, insertion at different sites would generate distinct bands. The rare differences detected by *Eco*RI would be expected if there are sites flanking the cloned fragment within the element . Based on the presence of reverse-transcriptase consensus sequences, all plants so far tested have been found to contain *copia*-like retrotransposons (Voytas et al. 1992). It will be interesting to determine if a *copia*-like reverse transcriptase from sorghum co-segregates with the fragments to which TXS1075 hybridizes.

Acknowledgements We thank Dr. Lei Zhang for assistance in DNA sequencing.

References

Aldrich PR, Doebley J, (1992) Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild *Sorghum bicolor*. Theor Appl Genet 85:293–302

Aldrich PR, Doebley J, Schertz KF, Stec A (1992) Patterns of allozyme variation in cultivated and wild *Sorghum bicolor*. Theor Appl Genet 85:451–460

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Bernatzky R, Tanksley SD (1989) Restriction fragments as molecular markers for germplasm evaluation and utilisation. In: Williams JT (ed) The use of plant genetic resources. Cambridge University Press, Cambridge, pp 353–363

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion J-P (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein-kinase family member. Cell 68:799–808

de Wet JMJ (1978) Systematics and evolution of *Sorghum* section *Sorghum* (Gramineae). Am J Bot 65:477–484

de Wet JMJ, Huckabay JP (1967) Origin of *Sorghum bicolor*. II. Distribution and domestication. Evolution 21:787-802

de Wet JMJ, Harlan JR, Price EG (1970) Origin of variability in the Spontanea complex of *Sorghum bicolor*. Am J Bot 57:704–707

Doggett H (1988) Sorghum, 2nd edn. John Wiley and Sons, New York, pp 5–69

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Bioch 123:6–13

Fraser NJ, Boyd Y, Craig I (1989) Isolation and characterization of a human variable copy-number tandem repeat at Xcen-Xp11.12. Genomics 5:144–148

Harlan JR, de Wet JMJ (1972) A simplified classification of cultivated sorghum. Crop Sci 12:172–1756

Helentjaris D, Slocum M, Wright S, Schaefer A, Nienhuis J (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. Theor Appl Genet 72:761–769

Hong-Tang, S, Dube K, Liang GH, Kung S (1990) Possible repetitive DNA markers for *Eusorghum* and *Parasorghum* and their potential use in examining phylogenetic hypotheses on the origin of *Sorghum* species. Genome 34:241–250

IBPGR (1984) Seed conservation. GeneBank handbook II. International Board for Plant Genetic Resources, Rome, Italy

Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, Baker E, Warren, ST, Schlessinger D, Sutherland DR (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science 252:1711–1713

Lewin B (1990) Genes IV. Cell Press, Cambridge, Massachusetts, pp 482-490

Liang GHL, Casady, AJ (1966) Quantitative presentation of the systematic relationship among twenty-one sorghum species. Crop Sci 6:76–79

Lubbers EL, Gill KS, Cox TS, Gill BS (1991) Variation of molecular markers among geographically-diverse accessions of *Triticum tauschii*. Genome 34:354–361

Maniatis T, Fritsch, EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Mann JA, Kimber CT, Miller FR (1983) The origin and early cultivation of sorghums in Africa. Bulletin no. 1454. Texas A and M University, College Station, Texas, USA

Morden WC, Doebley J, Schertz KF (1989) Allozyme variation in old world races of *Sorghum bicolor* (*Poaceae*). Amer J Bot 76:245–255

Morden WC, Doebley J, Schertz KF (1990) Allozyme variation among the spontaneous species of *Sorghum* section Sorghum (*Poaceae*). Theor Appl Genet 80:296–304

Murray MG, Thompson WF (1980) Rapid isolation of high-molecular-weight plant DNA. Nucleic Acids Res 8:4321–4325

Redman JB, Fenwick RJ Jr, Fu Y-H, Pizzuti A, Caskey CT (1993) Relationship between parental trinucleotide GCT repeat length and severity of myotonic dystrophy in offspring. JAMA 269: 1960–1965

Reed KC, Mann DA (1985) Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res 13:7207–7221

Sanger F, Couson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94:441–448

Sanger F, Nicklen S, Couson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schertz KF, Ritchey JM (1978) Cytoplasm-genic male-sterility systems in sorghum. Crop Sci 18:890–893

Voytas DF, Cummings MP, Konieczny A, Ausubel FW, Rodermel SR (1992) *Copia*-like retrotransposons are ubiquitous among plants. Proc Natl Acad Sci USA 89:7124–7128

White O, Soderlund C, Shanmugan P, Fields C (1992) Information contents and dinucleotide compositions of plant intron sequences vary with evolutionary origin. Plant Mol Biol 19:1057–1064

Xu GW, Magill CW, Schertz KF, Hart GE (1994) An RFLP linkage map of *Sorghum bicolor* (L.) Moench. Theor Appl Genet 89: 139–145